TUM – Courses
IN2045
Network Analysis – statistical and formal models and methods

Dr. Heiko Niedermayer, Cornelius Diekmann
(many slide sets from lecture Simulationstechnik by Prof. Alexander von Bodisco)

Chair for Network Architectures and Services

Department of Computer Science
Technische Universität München
http://www.net.in.tum.de

Technische Universität München

Course organization

- Old course IN2045 "Simulationstechnik"
  will be changed into
  "Netzwerkanalyse – statistische und formale Modelle und Methoden"

- A lot of topics and slides from the old course will be reused.
  - Most of them are from Alexander von Bodisco.
  - Other contributors to his slides will be mentioned on the slide set where applicable.

- There will be
  - Less: simulation, construction of simulators
  - More: measurement, models and modelling aspects
  - New: formal methods such as Higher-Order Logic, multi-disciplinary considerations
  - Tools: Python / Scipy / Numpy, Isabelle/HOL, ….

Course organization

- Lecturer
  - Dr. Heiko Niedermayer, niedermayer@net.in.tum.de
  - Cornelius Diekmann, diekmann@net.in.tum.de
  - Prof. Dr. Georg Carle, carle@net.in.tum.de
- Course
  - Lecture and Exercise share same time slots. Exercises will be announced.
  - Tuesday 16:00-18:00, Room: 03.07.023
  - Thursday 14:00-16:00, Room: 03.07.023
- ECTS:
  - 5 credits
- Exam:
  - Written exam or oral exam at the end of the semester.
- Course Material:
  - http://www.net.in.tum.de/de/lehre/ss14/vorlesungen/netzwerkanalyse-statistische-und-formale-modelle-und-methoden

Course organization

- Exercises:
  - Will not be corrected.
  - We want voluntary participation and presentation of solutions.

- Teaching Concepts:
  - Lecture, Exercises + more interactive methods will be applied in both

- Goal:
  - Get familiar with statistical issues (statistical significance)
  - Get familiar with models and modelling
  - Learn how to model systems
  - Learn how to evaluate different systems (formal/simulation/measurements)
  - Learn how to analyze results by using related tools and methods
  - Learn how to visualize results

⇨ **Prepare students for their BA/MA thesis**

Course Outline

Tools and methods will be introduced applicable and suitable in the particular modelling context.

Outline:

- Short Introduction to Modelling
- Discrete Deterministic Modelling
- Modelling of Systems with Randomness
- Advanced topics: e.g. Internet Science

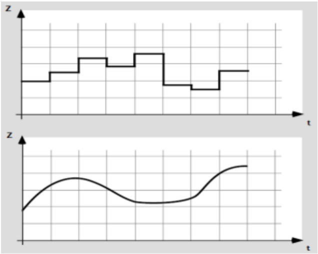
From the Old Course

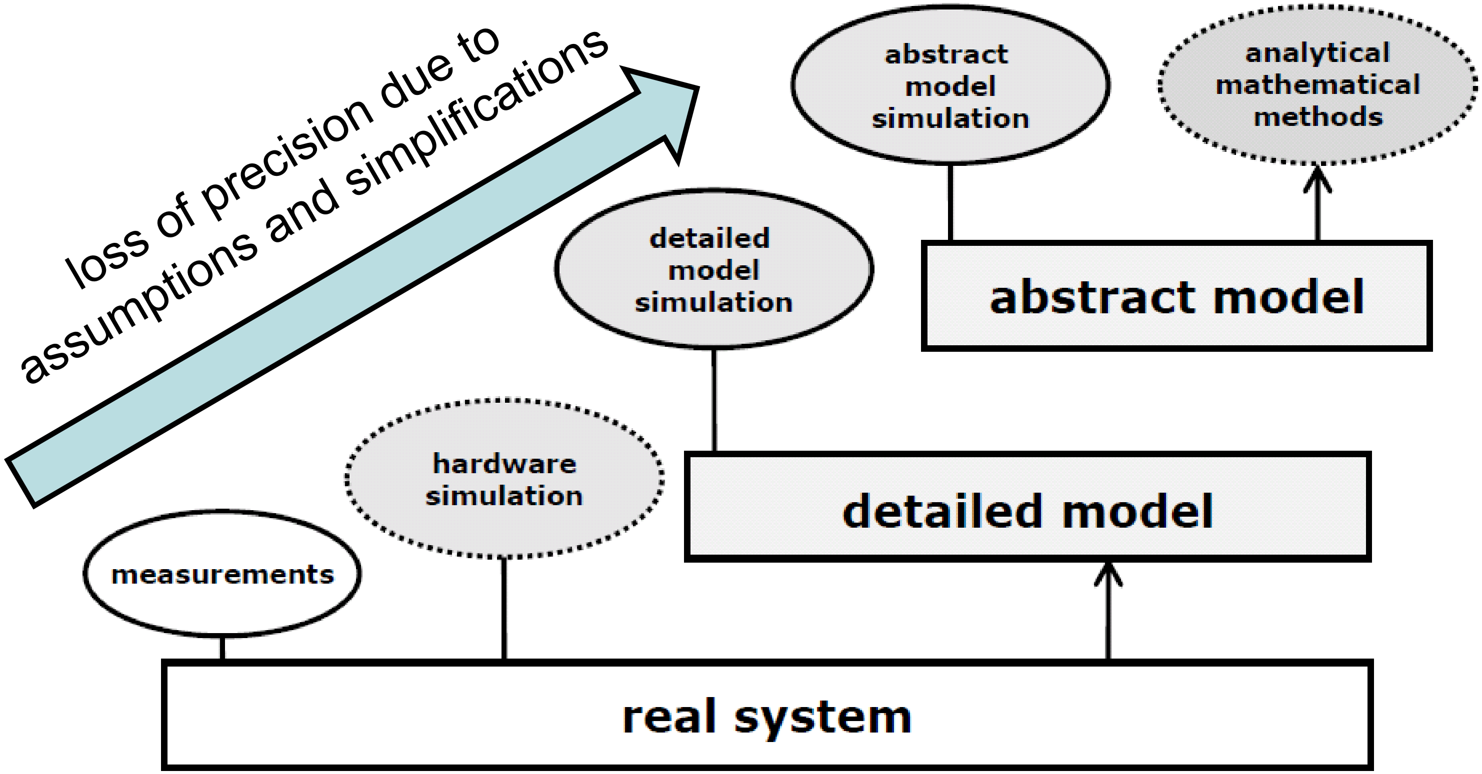
- In the following some slides on topics from the old course that we cover in either complete or (more often) reduced form.

Highlights from the Old Course

Simulation

- Simulation: What it is and when to use it
- Types of simulators
- Internals of discrete event simulators
- Continuations and co-routines

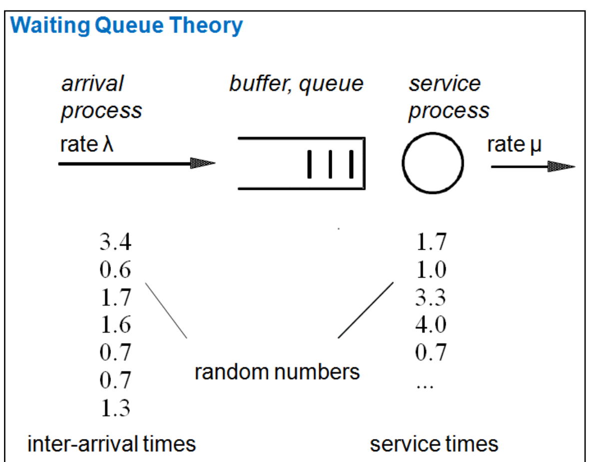


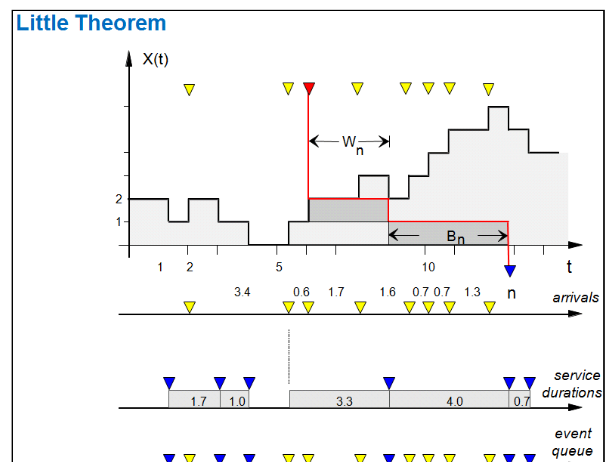


Highlights from the Old Course

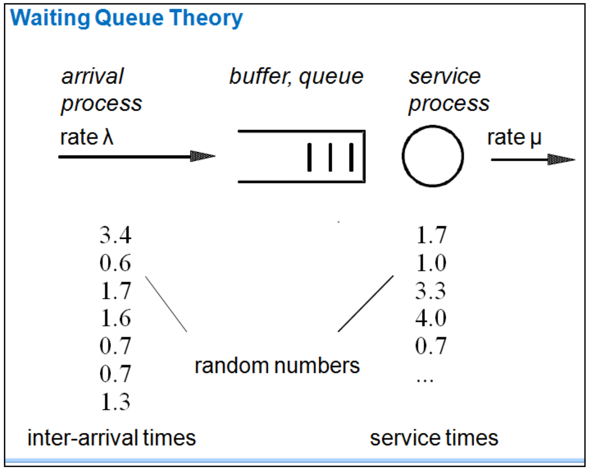
Statistics fundamentals

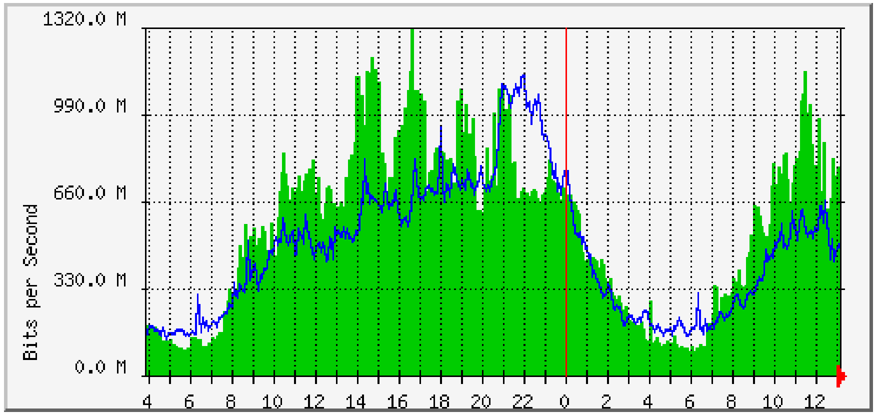
- Introduction to Waiting Queues
- Random Variable (RV), Discrete and Continuous RV
- Probability Space, Frequency Probability
- Distribution(discrete), Distribution Function(continuous)
- Probability Density Function, Cumulative Density Function
- Definitions: Expectation/Mean, Mode, Standard Deviation, Variance, Coefficient of Variation, p-percentile(quantile), Skewness, Scalability Issues, Covariance, Correlation, Autocorrelation Visualization of Correlation





Statistics fundamentals





Throughput MWN – Router - Garching

- Single high performance service process vs. multiple low performance service processes
- Impact for limited buffer size / storage capacity
- State / time dependent arrival process
- Performance parameters

Highlights from the Old Course

Random Numbers

- Random Variables:
  - Generation of Random Variables (RV)

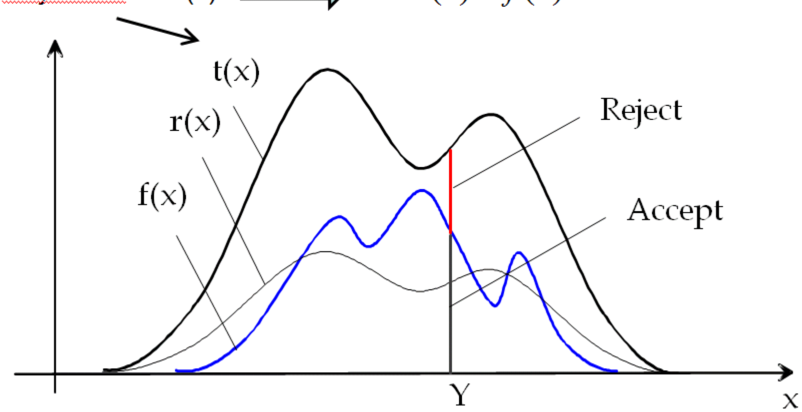
    Inversion, Composition, Convolution, Accept-Reject
  - Distributions and their Characteristics

    Uniform(continuous), Normal, Triangle, Lognormal, Exponential, Erlang-k, Gamma, Uniform(discrete), Bernoulli, Geom, Poisson, General Discrete

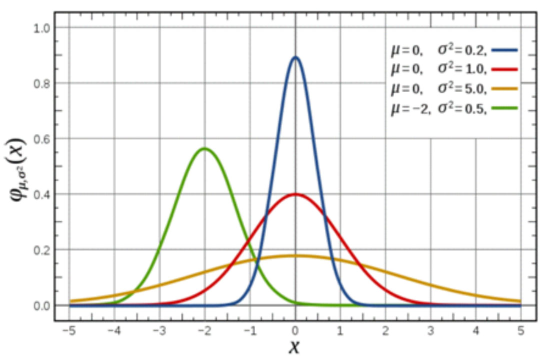
Accept-Reject

- Geometrical interpretation

  Y will be accepted if the point $(Y, U \cdot t(Y))$ falls under the curve *f*.
- The acceptance probability is high if *t(Y)-f(Y)* is small.
- Majorante von f(x) ⟹ $\forall x: t(x) \geq f(x)$



- **Normal distribution(3/3):**



Probability Density Function

Random Numbers

00101110101001101100010011101010100011

00101110101001101100010011101010100011

Random 

Autocorrelation Lag 4

Highlights from the Old Course

Evaluation of simulation results:

- Consistent Estimator, Unbiased Estimator, Variance of an Estimator, Bessel's Correction, Efficient Calculation
- Confidence Interval
  - Chebyshev
  - Central Limit Theorem
  - t-Distribution
- Evaluation and comparison of Simulation Results Replicate-Delete Method, Batch Means Method, Stationarity

Confidence interval according to the central limit theorem

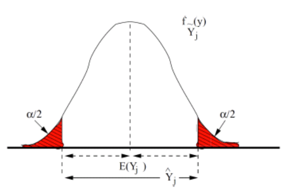
- Idea: The central limit theorem is still valid if σ^2 is replaced by $\tilde{S}^2$. Thus, it is possible to calculate the critical values out of the normal distribution.
- Recapitulate the “flipping of a coin example” with $\tilde{Y}$ representing the distribution of the estimator and Y being the distribution of the estimand. Then we can calculate the confidence interval as follows:

$$\Rightarrow \quad P[|Z| \geq \varepsilon] = P\left[\left|\frac{\tilde{Y} - E(Y)}{\tilde{S}/\sqrt{n}}\right| \geq \varepsilon\right] = \alpha$$

$$\Rightarrow \quad P[|\tilde{Y} - E(Y)| \geq \varepsilon \cdot \tilde{S}/\sqrt{n}] = \alpha$$

$$\Rightarrow \quad \boxed{\tilde{Y} \pm z_\alpha \cdot \tilde{S}/\sqrt{n}}$$

$\Rightarrow$ z_α is the $\alpha/2$ percentile of $N(0,1)$



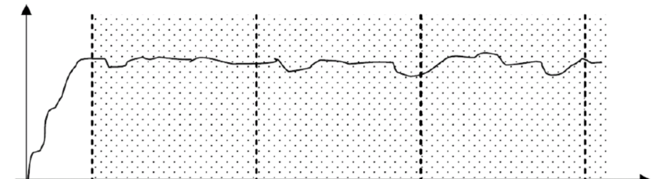
Picture taken from Buchholz

Batch-Means Method (LK 9.5.3)

- Estimate the duration of the transient phase
- Perform a long simulation run
- Remove the transient phase
- Divide the gathered results in n intervals of equal length (Batches) which hold m samples

$\Rightarrow$ Assure that the mean of subsequent batches is uncorrelated (calculate the empirical autocorrelation)

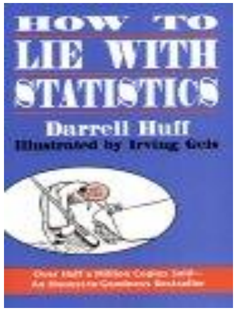
$\Rightarrow$ Number of batches $n \geq 10$ Batch size $m \geq 10 \cdot x$

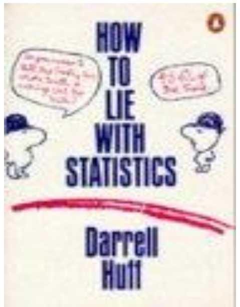


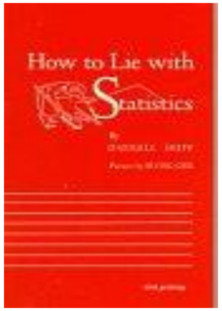
Highlights from the Old Course

How to Lie with Statistics:

- Lessons for Authors and Readers
- Examples and Discussion



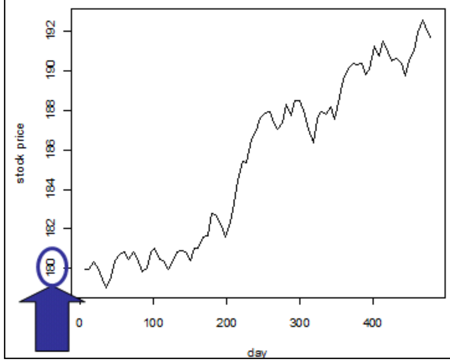


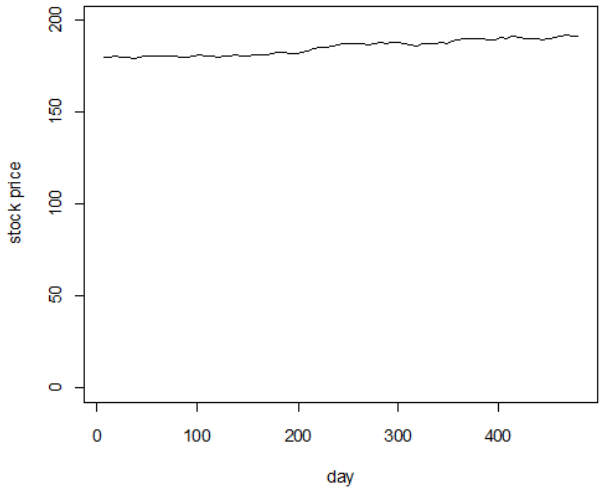


Scales can be misleading

- What really happened is shown here:

  We intuitively interpret a trend plot on a ratio





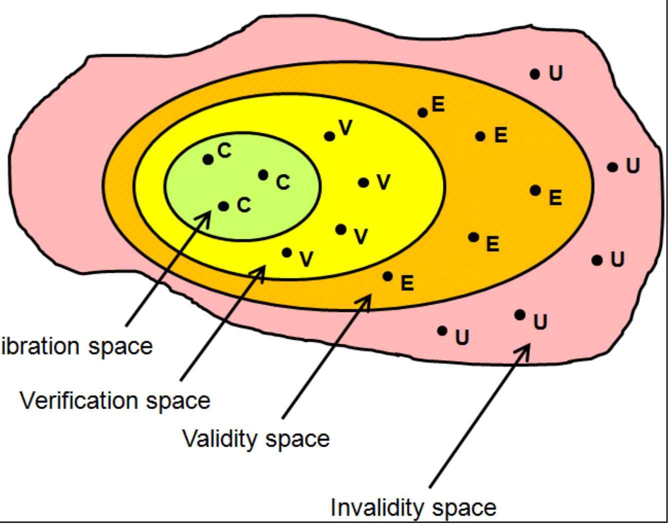
Highlights from the Old Course

Evaluation of simulation results:

- Model Validation:
  - Calibration, Overfitting
  - Structural Change, Parameter Change
  - Comparison of Confidence Intervals:

    Welsh, Law & Kelton

Calibration and Validation



Comparison of confidence intervals

Welch

- Estimators

$$\Rightarrow \quad \tilde{\mu}_R = \frac{1}{n}\cdot\sum_{i=1}^{n} V_{R_i} \qquad \tilde{S}_R^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(V_{R_i} - \tilde{\mu}_R\right)^2$$

$$\Rightarrow \quad \tilde{\mu}_S = \frac{1}{m}\cdot\sum_{i=1}^{m} V_{S_i} \qquad \tilde{S}_S^2 = \frac{1}{m-1}\sum_{i=1}^{m}\left(V_{S_i} - \tilde{\mu}_S\right)^2$$

- Difference of both samples are defined as follows:
  - $v_{RS\,i} = v_{Ri} - v_{Si}$
  - $v_{RS} = \{v_{RS_1}, v_{RS_2}, \ldots, v_{RS_n}\}$
  - $\tilde{\mu}_{RS} = \frac{1}{n}\cdot\sum_{i=1}^{n} V_{RS_i}$
  - $\tilde{S}_R^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(V_{R_i} - \tilde{\mu}_R\right)^2$

$$\Rightarrow \quad \boxed{\hat{\mu}_{RS} \pm t_{n-1,1-\alpha/2}\cdot \tilde{S}_{RS}/\sqrt{n}}$$

Both samples have to be statistically independent.

The samples must be of the same size. $m = n$

The variance of both samples must be equal. $Var(V_R) = Var(V_S)$

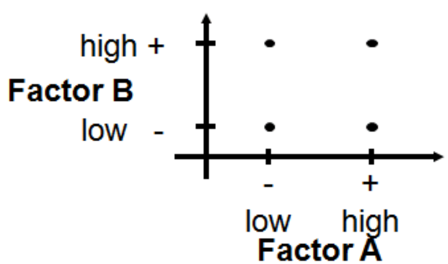
Highlights from the Old Course

Experiment planning:

- Hypothesis Testing
- Linear Regression
- Factorial Design

2^k factorial designs

- Example: 2 factors, i.e., a 2^2 design
- 4 design points:



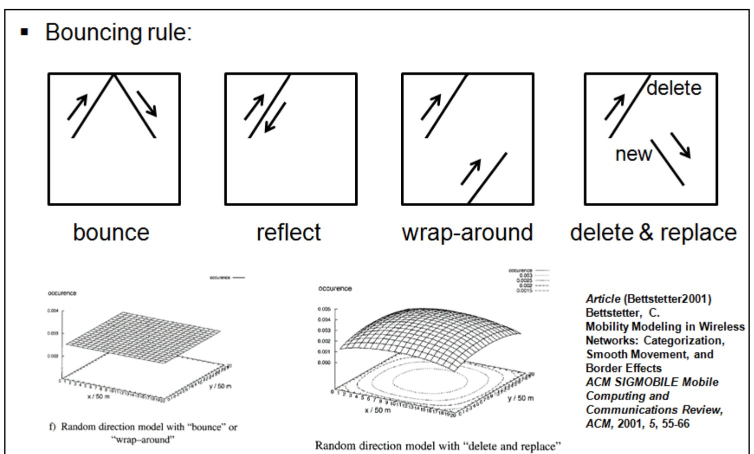
- Design matrix:

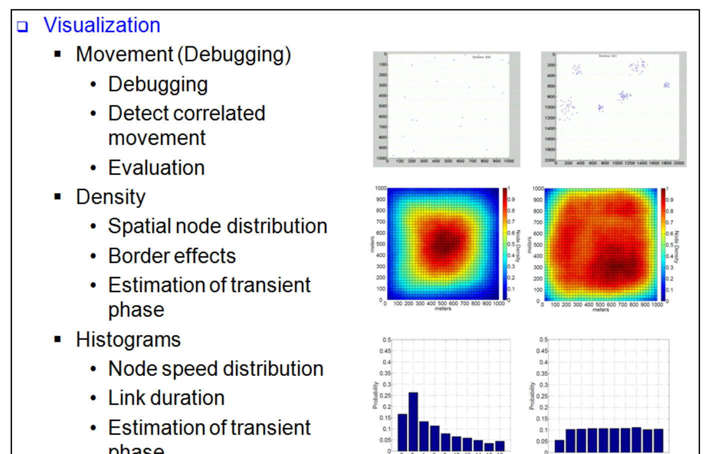
Run	Factor A	Factor B	Response
1	-	-	r_1
2	+	-	r_2
3	-	+	r_3
4	+	+	r_4

Highlights from the Old Course

Mobility:

- Mobility in General
  - Human Mobility Pattern
  - Visualization:
    - Density, Speed Histograms, Bouncing Rule, Obstacles
- Characteristics of Mobility Pattern:
  - Link Duration, Transient Phase, Node Distribution, Speed Distribution, Correlated Movement
- Synthetic Mobility Models:
  - Random Waypoint, Random Direction, Random Walk, Levi-Flight, Brownian Motion, Group Mobility

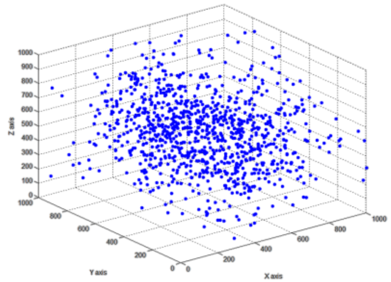


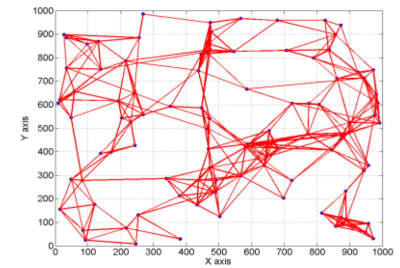


Highlights from the Old Course

Random Graphs:

- Graph Definition
- Generation of Random Graphs
- Probabilistic Model, Waxman Model
- Random Graphs with Predefined Characteristics
- Scale-free Graphs, Social-networks





- Scale-free networks – real-world examples:
  - **Social networks - Facebook:**

Facebook Friendships

Picture taken from http://www.opte.org

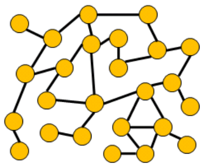
- **Scale-free graph:**

  A graph is called scale-free if its node degree k follows the power law.

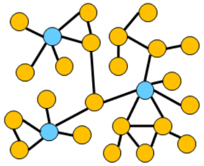
  $$P(k) = ck^{-\gamma}$$

  c and γ are constants. Typical range 0 < c < 1, 2 < γ < 3.

  - **Examples:**
    - Social networks
    - Collaboration networks
    - Computer networks
    - Disease transmission



Random Graph



Scale-free Graph